

MATH 104A FINAL PROJECT

Least Square Approximation and its Presence in Financial Mathematics

By: Kevin Payton, Grace Slone, Edwin Wang, Shenyang Zeng

March 18, 2024

Contents

1	Introduction	2
1.1	Problem Statement	2
1.2	Brief History and Overview	2
1.3	Overview of Financial Markets	3
1.4	Relevance of L.S.A in Finance	3
2	L.S.A Process	4
2.1	Detailed Explanation of the L.S.A Process	4
2.2	Proof of L.S.A	4
3	Code and Data Analysis	5
3.1	Overview of L.S.A Code	5
3.2	Example Data to Showcase L.S.A	7
3.3	Analysis of Data	8
4	Conclusion	9

1 Introduction

1.1 Problem Statement

This project investigates the use of the least square approximation to optimize returns while minimizing risk. The least square approximation method offers significant potential in achieving these objectives.

To begin, we will introduce some historical context of the l.s.a and discuss why it was created in the first place, as well as how it can be related to financial markets. We will then go through a detailed explanation and proof of the full process for determining your approximations. To finish things off, we will analyze the effectiveness of the l.s.a in doing the optimization outlined above through two different examples.

1.2 Brief History and Overview

The least square method is a form of mathematical regression analysis used to find the line of best fit for a given data set. Analysts can then use this line of best fit to determine if there is a relationship present between the data points, and make predictions based on these relationships. The least square method is widely used in modern finance and investing due to the importance of prediction and lowering risk.

The method of least squares approximation was discovered independently by Carl Friedrich Gauss and Adrien-Marie Legendre. Although Gauss had been using it since around 1795, Legendre was the first to publish a work using the method in 1805. Gauss released his version of the approximation four years later and is often credited with the discovery. (1)

The use of the least squares grew initially in the fields of astronomy and geodesy since scientists sought to determine ways to accurately navigate the oceans. A description of the behavior of celestial bodies was the key to allowing ships to sail the open seas. (6) One of the earliest uses of the l.s.a was to find the position of Ceres, the largest object in the Asteroid Belt between Mars and Jupiter. In the early 19th century, a newly discovered celestial object, later called Ceres, gained much attention from the scientific community due to a buzz about extra-terrestrial life. Unfortunately, the position of Ceres was lost a few months after it re-emerged from the glare of the Sun, and scientists had no way of determining its position with their limited data. Gauss was able to solve this problem and accurately determine the parameters of the orbit of Ceres using his l.s.a method. He was the only scientist to accurately predict the position of the asteroid, which catapulted both Gauss and the l.s.a method to fame. (7)

1.3 Overview of Financial Markets

A financial market refers to infrastructure or systems where goods or services are exchanged, typically for monetary gain. In a majority of financial markets, financial securities are often traded. Financial securities include valuable goods such as materials or metals, and instances of financial ownership such as stocks and bonds. These markets are necessary for capitalism, as they ensure money and assets are constantly moving and rarely stagnant. The importance of fluid money stems from the value of the asset itself, as the liquidity of capital ensures it does not lose value when exchanged.

One of the most active financial markets is the Foreign Exchange market, which involves individuals buying, selling, and predicting exchange rates of various foreign currencies (8). Similarly, stock markets involve companies selling shares, or partial business ownership, and investors predicting the value of these shares in the future. Commodity markets, on the other hand, involve the exchange of goods and raw materials for profit.

The failure or collapse of these markets can lead to increased levels of unemployment, economic recession, or even complete economic collapse. For instance, the stock market crash in 1929 was the direct result of stock prices rising to such heights that a majority of investors could not keep up. Thus, to prevent such events from happening, accurate predictive analysis of financial data is crucial.

1.4 Relevance of L.S.A in Finance

As mentioned in the previous section, an integral part of financial markets is prediction of value. Failing to accurately predict or recognize financial trends could result in a small scale incident such as an investor losing their savings, or could result in a much larger incident, such as economic collapse. Financial analysts arrive at such predictions through methods such as the Least Square Approximation.

The Least Square Approximation is thoroughly prevalent in many areas of finance since it allows researchers to notice correlations between variables. Noticing such trends and relationships provides more information and allows investors to make smarter investments. For instance, if there appears to be a negative trend in an investment that an individual has made, the investor may choose to sell that investment before its value decreases further. On the other hand, if there appears to be a positive trend in the sale of a certain raw material, an investor may choose to invest in it while its price is reasonable, eventually selling it for a much higher value than initially bought for.

2 L.S.A Process

2.1 Detailed Explanation of the L.S.A Process

Least Squares Analysis (LSA) is a mathematical method used to find the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve. In the context of financial analysis, LSA is often applied in linear regression to model the relationship between two variables. The process involves:

Model Specification: Defining the linear relationship $y=mx+b$, where y is the dependent variable (e.g., stock price), x is the independent variable (e.g., market index), m is the slope of the line, and b is the y-intercept.

Parameter Estimation: Calculating the values of m and b that minimize the sum of squared residuals between the actual data points and the predictions made by the linear model.

Model Evaluation: Assessing the model's fit by examining the residuals and calculating statistics like the Pearson correlation coefficient and R-squared value to quantify the strength and direction of the linear relationship.

2.2 Proof of L.S.A

Given a set of observations (x_i, y_i) for $i = 1, 2, \dots, N$, where x_i are the predictors (independent variables) and y_i are the responses (dependent variables), the goal of least squares regression is to find a linear relationship that best fits these data points. This relationship is represented as:

$$y_i = w^T x_i + \epsilon_i$$

where w is the weight vector that we want to estimate, and ϵ_i is the error term for each observation. The objective is to minimize the sum of squared errors (SSE), given by:

$$SSE = \sum_{i=1}^N (y_i - w^T x_i)^2$$

Matrix Formulation To solve this in a compact form, we can write the equations using matrices:

$$Y = XW + \epsilon$$

where Y is the vector of observed values, X is the matrix of predictors with each row representing a data point x_i^T , W is the vector of weights, and ϵ is the vector of errors. The SSE can then be

rewritten in matrix notation as:

$$SSE = (Y - XW)^T(Y - XW)$$

Derivation To find the minimum of SSE with respect to W , we set the derivative of SSE with respect to W to zero:

$$\frac{\partial SSE}{\partial W} = -2X^T(Y - XW) = 0$$

Solving for W , we get:

$$X^TY = X^TXW$$

$$W = (X^TX)^{-1}X^TY$$

This solution, $W = (X^TX)^{-1}X^TY$, is known as the normal equation for linear regression, which gives us the least squares estimates of the parameters that minimize the SSE.

Summary The least squares approximation method uses matrix algebra to find the weight vector W that minimizes the sum of squared differences between the observed and predicted values. The key step is to solve the normal equation, $W = (X^TX)^{-1}X^TY$, which yields the optimal parameters for the linear regression model.

3 Code and Data Analysis

3.1 Overview of L.S.A Code

The Least Squares Analysis (L.S.A) method is a fundamental approach in statistical modeling used to determine the best fit line that minimizes the sum of the squares of the residuals—the differences between observed and predicted values. In the context of financial mathematics, L.S.A is instrumental in examining the relationship between various financial indicators, such as stock performances compared to market indexes or commodity prices.

Our Python implementation leverages the ‘**numpy**’ library for numerical computations and ‘**matplotlib.pyplot**’ for visualizing the data alongside its line of best fit. The core of our L.S.A code comprises a function that calculates the slope and intercept of the line that best fits the data points in a linear relationship. This calculation is based on the least squares criterion, optimizing the accuracy of predictions made from the model.

The code for the l.s.a approximation is as follows:

```

1 def least_squares(x, y):
2     """
3     Calculates the least squares solution to a linear system.
4
5     Parameters:
6     x (array-like): The independent variable values.
7     y (array-like): The dependent variable values.
8
9     Returns:
10    tuple: Coefficients for the linear equation  $y = mx + b$ 
11    """
12    x = np.array(x)
13    y = np.array(y)
14    sum_x = np.sum(x)
15    sum_y = np.sum(y)
16    sum_xy = np.sum(x*y)
17    sum_x_squared = np.sum(x**2)
18    n = len(x)
19
20    m = (n*sum_xy - sum_x*sum_y) / (n*sum_x_squared - sum_x**2)
21    b = (sum_y - m*sum_x) / n
22
23    return m, b

```

The code for the plotting of data is as follows:

```

1 def plot_data_and_fit_with_correlation(x, y, x_label='X', y_label='Y'):
2     """
3     Plots the given x and y data along with the line of best fit and
4     displays the Pearson correlation coefficient.
5
6     Parameters:
7     x (array-like): The independent variable values.
8     y (array-like): The dependent variable values.
9     x_label (string): The x-axis label.
10    y_label (string): The y-axis label.
11    """
12    # Calculate line of best fit

```

```

13     m, b = least_squares(x, y)
14
15     # Calculate Pearson correlation coefficient
16     correlation_matrix = np.corrcoef(x, y)
17     correlation_coef = correlation_matrix[0, 1]
18
19     # Create line of best fit
20     x_fit = np.linspace(min(x), max(x), 1000)
21     y_fit = m*x_fit + b
22
23     # Plot data and line of best fit
24     plt.figure(figsize=(8, 6))
25     plt.scatter(x, y, color='blue', label='Data Points')
26     plt.plot(x_fit, y_fit, color='red', label=f'Best Fit: y = {m:.2f}x + {b:.2f}')
27     plt.xlabel(x_label)
28     plt.ylabel(y_label)
29     plt.title(f'Data Points, Line of Best Fit, and Correlation Coefficient:
    ↪ {correlation_coef:.2f}')
30     plt.legend()
31     plt.grid(True)
32     plt.show()

```

3.2 Example Data to Showcase L.S.A

To showcase our code, we decided to create a model comparing the price of a share in Microsoft, the largest company by market cap, and the S&P500, an index fund of the top 500 U.S.-based companies by market cap. The S&P500 allocates weight percentages based on how large a company's market cap is compared to all the others. Microsoft has the largest weight on this index fund and should enact the most change when its valuation fluctuates.

We also decided to create a model comparing the price of a share of Exxon Mobil Corp., a gas and oil company, and the price of gas in the U.S. Although the price of gas is much more regulated than the price of stocks, we would expect the share price to follow the price of gas fairly closely.

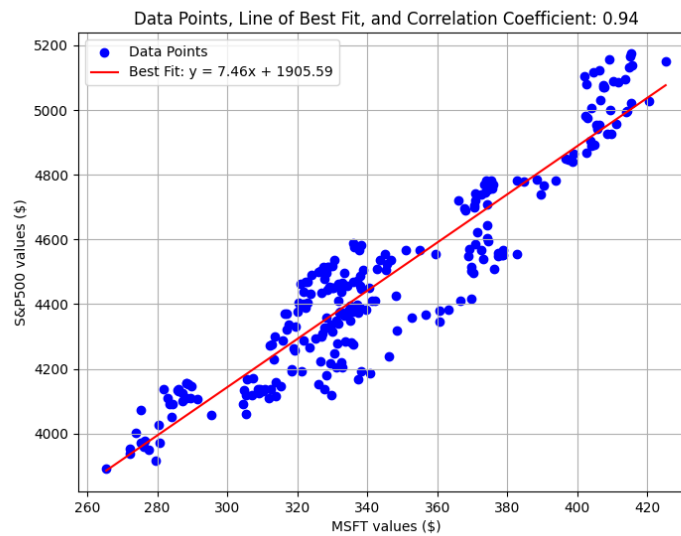


Figure 1: Comparison between the price of a MSFT share and the price of the S&P500. We can clearly see a linear relationship using the l.s.a method, and the correlation coefficient of 0.94 signifies a strong relationship.

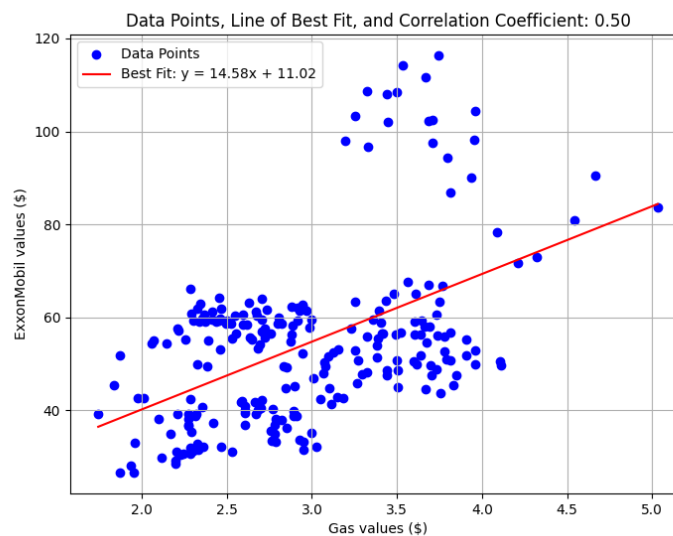


Figure 2: Comparison between the price of gas in the U.S. and the price of an ExxonMobil share. There is a linear relationship present, but due to the outliers and the correlation coefficient of 0.50, there is only a moderate relationship between the two.

3.3 Analysis of Data

Microsoft Corp (MSFT) vs. S&P 500 Index

Our analysis revealed a strong positive correlation between Microsoft's stock performance and

the S&P 500 index, as expected for a leading technology company in the stock market. The linear regression model, demonstrated through L.S.A, effectively captures this relationship, indicating that Microsoft's stock movements can be partially predicted by observing broader market trends. This insight is invaluable for investors seeking to align their portfolios with market dynamics.

ExxonMobil Corp (XOM) vs. Oil Prices

The relationship between ExxonMobil's stock prices and oil prices presented a more nuanced analysis. While conventional wisdom suggests that ExxonMobil's financial performance should directly correlate with oil prices, the analysis highlighted periods of divergence. Notably, during the COVID-19 pandemic, ExxonMobil's stock price experienced significant volatility, deviating from the rising trend in oil prices. This anomaly underscores the complexity of market behaviors and the influence of external factors on stock prices, challenging the predictive power of linear models.

Importance of L.S.A in Financial Modeling

The application of L.S.A in these case studies underscores the importance of statistical models in financial analysis. While the model provides a foundational understanding of the relationships between different financial indicators, it also highlights the limitations of linear regression in capturing the full spectrum of market dynamics. External factors, such as global pandemics, can significantly impact these relationships, demonstrating the need for a comprehensive approach to financial analysis that incorporates both quantitative models and qualitative assessments.

4 Conclusion

The exploration of the Least Square Approximation (L.S.A) method within the realm of financial mathematics, as detailed in this project, underscores its pivotal role in enhancing predictive accuracy and optimizing investment strategies. We have demonstrated its efficacy in modeling complex financial phenomena through a meticulous process involving the formulation, proof, and application of the L.S.A method.

The case studies involving Microsoft Corp (MSFT) versus the S&P 500 Index and ExxonMobil Corp (XOM) versus oil prices vividly illustrate the utility of L.S.A in discerning and leveraging linear relationships between market indicators. These analyses reveal that while L.S.A can significantly illuminate trends and correlations, the impact of external factors, such as global pandemics, can introduce volatility and deviations that challenge the predictive capability of linear models.

Our findings highlight the indispensable value of L.S.A in financial analytics, offering a robust tool for investors to navigate the intricacies of the market. However, the limitations indicate a

pressing need for integrating L.S.A with more sophisticated models and qualitative assessments to better account for the unpredictable dynamics of financial markets.

The continuous evolution of financial instruments and market conditions calls for further research and development of advanced statistical methods. Integrating L.S.A with machine learning algorithms and big data analytics presents a promising frontier for enhancing predictive modeling and risk management strategies.

In conclusion, the journey through the conceptual framework, empirical analysis, and application of the Least Square Approximation method in financial mathematics has validated its significance and charted a path for future explorations aimed at refining investment decisions and fostering financial innovation.